

CLAIMS

Sub
A1 1. A method for processing a document including an internal structure made up of a plurality of elements, comprising:

a characteristic information extracting step of extracting characteristics of said document; and

a document classification step of classifying each document into a plurality of classification items making up a classification model depending on the degree of interrelation between the characteristic information of the document extracted by the characteristic information extracting step and the classification-item-based characteristic information.

2. The document processing method according to claim 1 further comprising a step of receiving a plurality of documents; said characteristic information extracting step extracting the characteristic information representing characteristics of each document received in said receiving step.

3. The document processing method according to claim 1 wherein a weight is imparted to each element based on the internal structure of said document; said characteristic information extracting step extracting an element having the weight larger than a pre-set value.

4. The document processing method according to claim 1 wherein said document classification step classifies the documents using a classification model prepared based on the classification operation of classifying the plural documents into classification

items.

5. The document processing method according to claim 1 wherein said document classification step calculates the degree of interrelation between the characteristic information of the document extracted in said characteristic information extracting step and the characteristic information of said classification items and classifies a document in a classification item with the maximum degree of interrelation when said degree of interrelation exceeds a threshold value.

SW3
AL 6. The document processing method according to claim 1 wherein the degree of interrelation of the characteristic information of the document extracted in said characteristic information extracting step and the characteristic information of said classification items is the linear linkage of the number of overlapping ones of proper nouns contained in the document characteristic information and of proper nouns contained in the characteristic information of said classification items and the degree of interrelation between the semantics other than the proper nouns contained in the document characteristic information and the semantics other than the proper nouns contained in said classification items.

7. The document processing method according to claim 6 wherein in obtaining the degree of interrelation of semantics, each semantics is weighted based on the structure of reciprocal referencing among the semantics, and a difference of a weight, imparted to one semantics, as a dependent variable, is divided by a difference of a weight, imparted to the other semantics, as an independent variable, with the resulting quotient

8. The document processing method according to claim 7 wherein the relation of reciprocal referencing among the semantics is prepared using a dictionary which, for each semantics, referencing is made to other semantics.

a weighting step for weighting each semantics based on the structure of reciprocal referencing among the semantics; and

10. The document processing method according to claim 9 wherein said weighting step imparts weight based on the structure of reciprocal referencing obtained on organizing the reciprocal referencing of semantics using a dictionary which, for each semantics, referencing is made to other semantics.

12. An apparatus for processing a document including an internal structure made up

characteristic information extracting means for extracting characteristics of said document; and

13. An apparatus for processing a document of an internal structure made up of plural semantics, said method comprising:

degree of interrelation calculating means for dividing a difference of a weight, imparted to one semantics, as a dependent variable, by a difference of a weight, imparted to the other semantics, as an independent variable, with the resulting quotient being the degree of interrelation between the one and the other semantics.

said document processing program including

a characteristic information extracting step of extracting characteristics of said document; and

